4 Nonstationary AR Process and Testing for Unit Root

- **Trend and Random Walk**

The general solution to a linear stochastic difference equation: $y_t = trend + stationary$ component + noise.

"**Trend**": permanent or nondecaying component of a time series. The trend has a permanent effect on a series.

"Trend stationary": $y_t = y_0 + a_0 t + A(L)\varepsilon_t$, where $A(L)\varepsilon_t$ is a stationary component of y_t.

"Not trend stationary": $y_t = y_0 + \sum_{i=1}^{t} \varepsilon_i + a_0 t$. $\sum_{i=1}^{t} \varepsilon_i$ is a stochastic trend component of y_t. Each ε_i shock has a permanent change in the conditional mean of $\{y_t\}$.

Random walk model:

$$y_t = y_0 + \sum_{i=1}^{t} \varepsilon_i$$

or

$$y_t = y_{t-1} + \varepsilon_t,$$

where ε_t is a white-noise. Nonstationary and Difference stationary

$$\begin{aligned} Ey_t &= Ey_{t-s} = y_0, \\ Var(y_t) &= t\sigma^2,\ Var(y_{t-s}) = (t-s)\sigma^2, \\ E_t y_{t+1} &= E_t y_{t+s} = y_t,\ s > 0. \end{aligned}$$

An ε_i shock ($i < t$) has a permanent effect on y_t and hence the forecasts for y_{t+s}. Since

$$Cov(y_t, y_{t-s}) = Cov\left(\sum_{i=1}^{t} \varepsilon_i, \sum_{i=1}^{t-s} \varepsilon_i\right) = (t-s)\sigma^2,$$

$$\rho_s = \frac{(t-s)\sigma^2}{\sqrt{t\sigma^2}\sqrt{(t-s)\sigma^2}} = (1 - s/t)^{1/2},$$

as s increases, ρ_s will decline, and hence the ACF for a random walk process will show a slight tendency to decay. It is impossible to use the ACF to distinguish between a unit root process and a stationary process with a near-unit root.

Random walk plus drift model:

$$y_t = y_0 + \sum_{i=1}^{t} \varepsilon_i + a_0 t$$

or

$$y_t = a_0 + y_{t-1} + \varepsilon_t,$$

where ε_t is a white-noise. Nonstationary: a linear deterministic trend $a_0 t$ and a stochastic trend $\sum_{i=1}^t \varepsilon_i$. It is difference-stationary.

$$E_t y_{t+s} = y_t + a_0 s.$$

Random walk plus noise model:

$$y_t = y_0 + \sum_{i=1}^t \varepsilon_i + \eta_t$$

or

$$y_t = y_{t-1} + \varepsilon_t + \Delta\eta_t,$$

where η_t is a white-noise with variance σ_η^2 and independent of ε_{t-s} for all t and s; $\eta_0 = 0$. It is Nonstationary but Difference-stationary.

$$\begin{aligned} Ey_t &= Ey_{t-s} = y_0, \\ Var(y_t) &= t\sigma^2 + \sigma_\eta^2,\ Var(y_{t-s}) = (t-s)\sigma^2 + \sigma_\eta^2, \\ E_t y_{t+1} &= E_t y_{t+s} = y_t,\ s > 0. \end{aligned}$$

The shock ε_t has a permanent effect on y_t and hence the forecasts for y_{t+s}. However, the noise η_t only affects y_t but not the subsequent values y_{t+s}. Since

$$Cov(y_t, y_{t-s}) = Cov\left(\sum_{i=1}^t \varepsilon_i + \eta_t, \sum_{i=1}^{t-s} \varepsilon_i + \eta_{t-s}\right) = (t-s)\sigma^2,$$

$$\rho_s = \frac{(t-s)\sigma^2}{\sqrt{t\sigma^2 + \sigma_\eta^2}\sqrt{(t-s)\sigma^2 + \sigma_\eta^2}},$$

the ACF is always smaller than that for the pure random walk model above. The noise is only to increase the variance of $\{y_t\}$ without affecting its long-run behavior. The model is nothing more than the random walk model with a purely temporary component added.

Trend plus noise model:

$$y_t = y_0 + a_0 t + \sum_{i=1}^t \varepsilon_i + \eta_t$$

or

$$y_t = a_0 + y_{t-1} + \varepsilon_t + \Delta\eta_t,$$

where η_t is a white-noise process with variance σ_η^2 and $E\varepsilon_t\eta_{t-s} = 0$ for all t and s. Nonstationary: a linear deterministic trend $a_0 t$, a stochastic trend $\sum_{i=1}^t \varepsilon_i$, and a pure white-noise η_t. It is difference-stationary.

General trend plus irregular model:

$$y_t = y_0 + a_0 t + \sum_{i=1}^{t} \varepsilon_i + A(L)\eta_t$$

or

$$y_t = a_0 + y_{t-1} + \varepsilon_t + A(L)\Delta\eta_t,$$

where $A(L)\eta_t$ is a stationary process. Nonstationary: a linear deterministic trend $a_0 t$, a stochastic trend $\sum_{i=1}^t \varepsilon_i$, and a stationary component $A(L)\eta_t$. Shocks to a stationary series are necessarily temporary; the effects of the irregular component will dissipate and do not affect its long-run mean level. But the trend components will determine the trend of the y_t process.

- **How to remove the trend?**

Two methods: Differencing and detrending.

1) Differencing:

- 1. Random walk plus drift model:

$$y_t = y_0 + \sum_{i=1}^{t} \varepsilon_i + a_0 t$$

or $\Delta y_t = a_0 + \varepsilon_t$, where ε_t is a white-noise. Stationarity of $\{\Delta y_t\}$:

$$\begin{aligned} E\Delta y_t &= a_0 \\ Var(\Delta y_t) &= \sigma^2 \\ Cov(\Delta y_t, \Delta y_{t-s}) &= 0, \; s > 0. \end{aligned}$$

2. Random walk plus noise model:

$$y_t = y_0 + \sum_{i=1}^{t} \varepsilon_i + \eta_t$$

or $\Delta y_t = \varepsilon_t + \Delta\eta_t$, where η_t is a white-noise with variance σ_η^2 and independent of ε_{t-s} for all t and s, and $\eta_0 = 0$. Stationarity of $\{\Delta y_t\}$:

$$\begin{aligned} E\Delta y_t &= 0 \\ Var(\Delta y_t) &= \sigma^2 + 2\sigma_\eta^2 \\ Cov(\Delta y_t, \Delta y_{t-1}) &= -\sigma_\eta^2 \\ Cov(\Delta y_t, \Delta y_{t-s}) &= 0,\ s > 1. \end{aligned}$$

Note that

$$-0.5 \le \rho_1 = \frac{Cov(\Delta y_t, \Delta y_{t-1})}{Var(\Delta y_t)} = \frac{-\sigma_\eta^2}{\sigma^2 + 2\sigma_\eta^2} \le 0$$

$$\rho_s = 0,\ s > 1.$$

$\{\Delta y_t\}$ acts exactly as an MA(1) process, and hence $\{y_t\}$ acts as $ARIMA(0,1,1)$.

3. Similarly, for the trend plus noise model,

$$y_t = y_0 + a_0 t + \sum_{i=1}^{t} \varepsilon_i + \eta_t$$

or $\Delta y_t = a_0 + \varepsilon_t + \Delta\eta_t$, where η_t is a white-noise process with variance σ_η^2 and $E\varepsilon_t\eta_{t-s} = 0$ for all t and s, $\{y_t\}$ acts as an $ARIMA(0,1,1)$ process.

4. Suppose that $A(L)y_t = B(L)\varepsilon_t$, where $A(L)$ and $B(L)$ are polynomials of order p and q, respectively. Assume also that $A(L)$ has a single unit root and that $B(L)$ has all roots outside the unit circle. Then $A(L) = (1 - L)A^*(L)$, where $A^*(L)$ is a polynomial of order $p-1$ whose roots are all outside of the unit circle. Hence, $(1 - L)A^*(L) = B(L)\varepsilon_t$ or $A^*(L)y_t^* = B(L)\varepsilon_t$, where $y_t^* = \Delta y_t$, and $\{y_t^*\}$ is stationary and $y_t \sim ARIMA(p-1, 1, q)$. If $A(L)$ has only two unit roots, the same argument deduces that $\{\Delta^2 y_t\}$ is stationary and $y_t \sim ARIMA(p-2, 2, q)$. Generally, suppose that $\{y_t\}$ is an $ARIMA(p, d, q)$ process (with d unit roots), then $\{\Delta^d y_t\}$ is stationary and $\{\Delta^d y_t\}$ is an $ARMA(p, q)$ process. In this case, $\{y_t\}$ is called "integrated of order d", i.e. $\{y_t\}$ is an $I(d)$ process or denoted as $y_t \sim I(d)$.

5. Trend plus noise model: $y_t = y_0 + a_1 t + \varepsilon_t$ or $\Delta y_t = a_1 + \varepsilon_t - \varepsilon_{t-1}$. $\{\Delta y_t\}$ is not well-behaved since it is not invertible, i.e. it cannot be expressed in the form of an autoregressive process. Hence, differencing does not work well here.

2) Detrending:

For the deterministic trend plus noise model: $y_t = a_0 + a_1 t + a_2 t^2 + \cdots + a_n t^n + \varepsilon_t$, applying detrending, i.e. regressing $\{y_t\}$ on the deterministic polynomial time trend

$(a_0+a_1t+a_2t^2+\cdots+a_nt^n)$ and subtracting the estimated values of y_t from the observed series y_t will yield estimated values $\{e_t\}$ of the $\{\varepsilon_t\}$ series, where $\{e_t\}$ is stationary. The degree of the polynomial can be determined by t-test, F-test, the AIC or SBC. (First try a larger n, apply t-test to the last coefficient, ...). The stationary detrended process can then be modeled using the traditional methods (ARMA estimation).

- **Difference or detrend?** Inappropriate method to eliminate trend will lead to a serious problem. (1) First-differencing the TS (trend-stationary) process will introduce a noninvertible unit root process into the MA component of the model. Examine the example

  $$A(L)y_t = a_0 + a_1t + B(L)\varepsilon_t.$$

  Detrending yields a satationary and invertible ARMA process, but first-difference deduces that

  $$A(L)\Delta y_t = a_1 + (1-L)B(L)\varepsilon_t$$

  of which the second term $(1-L)B(L)\varepsilon_t$ makes $A(L)\Delta y_t$ noninvertible. (2) Detrending a DS process may not eliminate all the trend components (the stochastic component of the trend). Study the general trend plus irregular model

  $$y_t = y_0 + a_0t + \sum_{i=1}^{t}\varepsilon_i + A(L)\eta_t.$$

- **Why nonstationary variables lead to spurious regression (high R^2,significant t-test, but no economic meaning)?** (the necessity or the importance of the unit root test). Examine the following example:

  $$y_t = a_0 + a_1z_t + e_t,$$

  where $\{y_t\}$ and $\{z_t\}$ are two independent random walk processes, i.e.

  $$\begin{aligned} y_t &= y_{t-1} + \varepsilon_{yt} \\ z_t &= z_{t-1} + \varepsilon_{zt} \end{aligned}$$

  with two independent white-noise processes ε_{yt} and ε_{zt}. Any relationship between y_t and z_t is meaningless since $\{y_t\}$ and $\{z_t\}$ are independent. But OLS often gives high R^2 and significant t-test for the coefficients. Why? The error term e_t in the regression equation is

  $$\begin{aligned} e_t &= y_t - a_0 - a_1z_t \\ &= -a_0 + \sum_{i=1}^{t}\varepsilon_{yi} - a_1\sum_{i=1}^{t}\varepsilon_{zi}. \end{aligned}$$

Since $Var(e_t) = t(\sigma^2_{\varepsilon_y} + a_1^2\sigma^2_{\varepsilon_z})$ becomes infinitely large as t increases, and $E_t e_{t+i} = e_t$ for $i \geq 0$, the t-test, F-test or R^2 values are unrealiable. Under the null: $a_1 = 0$, $y_t = a_0 + e_t$ and hence $\{e_t\} \sim I(1)$. This is inconsistent with the distributional theory in OLS, which requires that the error term be a white-noise. Phillips (1986) shows that, the larger the sample, the more likely to falsely reject the null (i.e. the more significant for the test in OLS).

- **Recall:**

If $\{\Delta y_t\}$ is a stationary process, then the process $\{y_t\}$ is called **a unit root process,** where $\Delta y_t = y_t - y_{t-1}$. Specially, if $\{\varepsilon_t\} \sim i.i.d.(0, \sigma^2)$, $\sigma^2 < \infty$, then the process $\{y_t\}$ with $y_t - y_{t-1} = \varepsilon_t$ is called **a random walk process**.

CLT for a Martingale Difference Sequence: Let $\{Y_t\}$ be a scalar martingale difference sequence with $\bar{Y}_T = (1/T)\sum_{t=1}^T Y_t$. Suppose that (a) $E(Y_t^2) = \sigma_t^2 > 0$ with $(1/T)\sum_{t=1}^T \sigma_t^2 \to \sigma^2 > 0$; (b) $E|Y_t|^r < \infty$ for some r>2 and all t; (c) $(1/T)\sum_{t=1}^T Y_t^2 \to \sigma^2$ in probability. Then

$$\frac{1}{\sqrt{T}}\sum_{t=1}^T Y_t \to N(0, \sigma^2). \qquad \blacksquare$$

Study the AR(1) model $y_t = a_1 y_{t-1} + u_t$ with $|a_1| < 1$, where $\{u_t\}$ is i.i.d$(0, \sigma^2)$. The OLS estimator $\hat{a}_1$ is

$$\hat{a}_1 = a_1 + \left(\sum_{t=1}^T y_{t-1}^2\right)^{-1} \sum_{t=1}^T u_t y_{t-1}.$$

This estimator is biased ($E\hat{a}_1 \neq a_1$) since u_t may not be independent of $y_t, y_{t+1}, \cdots, y_T$, but, as $T \to \infty$,

$$\sqrt{T}(\hat{a}_1 - a_1) \to N(0, 1 - a_1^2). \qquad \blacksquare$$

- **Winner Process:**

Now study

$$y_t = a_1 y_{t-1} + \varepsilon_t, \ a_1 = 1,$$

where $\varepsilon_t \sim i.i.d.N(0, \sigma^2)$.The OLS estimator for a_1 satisfies

$$T(\hat{a}_1 - 1) = \frac{(1/T)\sum_{t=1}^T \varepsilon_t y_{t-1}}{(1/T^2)\sum_{t=1}^T y_{t-1}^2},$$

which is not asymptotically normally distributed. We will show that the asymptotic distribution of $T(\hat{a}_1 - 1)$ is related with Winner process.

Winner Process (Standard Brownian Motion) $W(\cdot)$ is defined as a continuous-time stochastic process, associating each date $t \in [0,1]$ with the scalar $W(t)$ such that

(i) $W(0) = 0$;

(ii) For any dates $0 \le t_1 < t_2 < \cdots < t_k = 1$, the changes

$$[W(t_2) - W(t_1)], \quad [W(t_3) - W(t_2)], \quad \cdots, \quad [W(t_k) - W(t_{k-1})]$$

are independent random variables;

(iii) For any $0 \le t < s \le 1$, $W(s) - W(t) \sim N(0, s-t)$.

Notes:

1) For any $t \in [0,1]$, select $dt > 0$ such that $0 < t - dt < 1$. $W(t) - W(t-dt) = \eta_t \sim N(0, dt)$ or

$$W(t) = W(t-dt) + \eta_t \quad \text{with } \eta_t \sim N(0, dt)$$

is a random walk with the time-step dt.

2) Other continuous-time processes can be generated from standard Brownian motion. For example, $B(t) \equiv \sigma W(t) \sim N(0, \sigma^2 t)$, which is called as Brownian motion with variance σ^2; $Z(t) = W^2(t) \sim t\chi^2(1)$.

3) $W(t)$ is continuous but not differentiable in $t \in [0,1]$, where the distance of $W(t_1)$ and $W(t_2)$ is defined as

$$d(t_1, t_2) = \sqrt{E\left(W(t_1) - W(t_2)\right)^2}$$

for any $t_1, t_2 \in [0,1]$ and $t_2 > t_1$. The reasons are that, for any $t_0 \in [0,1], dt > 0, t_0 + dt \in [0,1]$, $W(t_0 + dt) - W(t_0) \sim N(0, dt)$, and hence

$$\begin{aligned} d(t_0, t_0 + dt) &= \sqrt{E\left(W(t_0 + dt) - W(t_0)\right)^2} \\ &= \sqrt{Var\left(W(t_0 + dt) - W(t_0)\right)} \\ &= \sqrt{dt} \to 0, \text{ as } dt \to 0 \end{aligned}$$

and

$$\lim_{dt \to 0^+} \frac{d(t_0, t_0 + dt)}{dt} = \lim_{dt \to 0^+} \frac{\sqrt{dt}}{dt} = \infty.$$

- **The Functional Central Limit Theorem**

If $\varepsilon_t \sim i.i.d.(0, \sigma^2), \sigma^2 < \infty$, then $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_t \to N(0, \sigma^2)$ as $T \to \infty$. This is the Lindeberg-Levy CLT. Also, as $T \to \infty$,

$$\frac{1}{\sqrt{[T/2]}} \sum_{t=1}^{[T/2]} \varepsilon_t \to N(0, \sigma^2).$$

Generally, for any $r \in (0, 1)$, as $T \to \infty$,

$$\frac{1}{\sqrt{[Tr]}} \sum_{t=1}^{[Tr]} \varepsilon_t \to N(0, \sigma^2).$$

Define

$$X_T(r) \equiv \frac{1}{T} \sum_{t=1}^{[Tr]} \varepsilon_t.$$

It is easily deduced that, for any given realization,

$$X_T(r) = \begin{cases} 0, & 0 \le r < 1/T \\ \varepsilon_1/T, & 1/T \le r < 2/T \\ (\varepsilon_1 + \varepsilon_2)/T, & 2/T \le r < 3/T \\ \vdots \\ (\varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_{T-1})/T, & (T-1)/T \le r < 1 \\ (\varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_T)/T, & r = 1 \end{cases}$$

is a step function in r. Since

$$\sqrt{T} X_T(r) \equiv \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} \varepsilon_t = \frac{\sqrt{[Tr]}}{\sqrt{T}} \frac{1}{\sqrt{[Tr]}} \sum_{t=1}^{[Tr]} \varepsilon_t$$

and $\lim_{T \to \infty} \frac{\sqrt{[Tr]}}{\sqrt{T}} \to \sqrt{r}$, we have

$$\sqrt{T} X_T(r) \equiv \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} \varepsilon_t \to N(0, r\sigma^2) \tag{1}$$

or

$$\sqrt{T} X_T(r)/\sigma \equiv \frac{1}{\sqrt{T}\sigma} \sum_{t=1}^{[Tr]} \varepsilon_t \to N(0, r).$$

For any $1 > r_2 > r_1 > 0$,

$$\sqrt{T} \left(X_T(r_2) - X_T(r_1) \right) / \sigma \to N(0, r_2 - r_1),$$

and it is independent of $\sqrt{T}X_T(r)/\sigma$, provided that $r < r_1$. Therefore, the set of functions $\{\sqrt{T}X_T(\cdot)/\sigma\}_{T=1}^{\infty}$ has an asymptotic distribution:

$$\sqrt{T}X_T(\cdot)/\sigma \rightarrow W(\cdot).$$

Note that here $X_T(\cdot)$ is a random function and $X_T(r)$ is a random variable, which is the value of the function $X_T(\cdot)$ at date r. The limiting distribution of $\sqrt{T}X_T(r)$ is the same as the distribution of the Wiener process $B(r) = \sigma W(r)$ since $W(r) \sim N(0, r)$. Specially,

$$\sqrt{T}X_T(1)/\sigma \equiv \frac{1}{\sqrt{T}\sigma}\sum_{t=1}^{T}\varepsilon_t \rightarrow W(1) \sim N(0,1).$$

- **Continuous Mapping Theorem:** If $S_T(\cdot) \rightarrow S(\cdot)$ in distribution and $g(\cdot)$ is a continuous function, then $g(S_T(\cdot)) \rightarrow g(S(\cdot))$ in distribution.

For example, since $\sqrt{T}X_T(\cdot)/\sigma \rightarrow W(\cdot)$, we have $\sqrt{T}X_T(\cdot) \rightarrow \sigma W(\cdot)$ and $\sqrt{T}X_T(r) \sim N(0, \sigma^2 r)$; $S_T(\cdot) \equiv \left(\sqrt{T}X_T(\cdot)\right)^2 \rightarrow \sigma^2 W^2(\cdot)$; $\int_0^1 \sqrt{T}X_T(r)dr \rightarrow \sigma \int_0^1 W(r)dr$.

- **Stochastic Integral**: $\int_0^1 f(r)dW(r)$ Definition and properties.

$$\int_0^1 dW = W(1) \sim N(0,1)$$

$$\int_0^1 W(t)dW(t) = \frac{1}{2}\left[W^2(1) - 1\right] \sim \frac{1}{2}\left[\chi^2(1) - 1\right]$$

$$\int_0^1 rdW(r) \sim N(0, 1/3)$$

- **Applications to Random Walk Processes:**

Theorem 1 Suppose that $y_t = y_{t-1} + \varepsilon_t$, where $\varepsilon_t \sim i.i.d.N(0, \sigma^2)$. Then as $T \rightarrow \infty$,

$$T^{-1/2}\sum_{t=1}^{T}\varepsilon_t \rightarrow \sigma W(1) \tag{2}$$

$$T^{-1}\sum_{t=1}^{T} y_{t-1}\varepsilon_t \rightarrow \frac{1}{2}\sigma^2(W^2(1) - 1) \tag{3}$$

$$T^{-3/2}\sum_{t=1}^{T} t\varepsilon_t \rightarrow \sigma W(1) - \sigma\int_0^1 W(r)dr \tag{4}$$

$$T^{-3/2}\sum_{t=1}^{T} y_{t-1} \to \sigma \int_0^1 W(r)dr \tag{5}$$

$$T^{-5/2}\sum_{t=1}^{T} t y_{t-1} \to \sigma \int_0^1 rW(r)dr \tag{6}$$

$$T^{-2}\sum_{t=1}^{T} y_{t-1}^2 \to \sigma^2 \int_0^1 W^2(r)dr. \tag{7}$$

Proof: (2) is obvious. For (3), note that $y_t^2 = (y_{t-1}+\varepsilon_t)^2 = y_{t-1}^2+\varepsilon_t^2+2\varepsilon_t y_{t-1}$. Therefore,

$$\begin{aligned}
T^{-1}\sum_{t=1}^{T} \varepsilon_t y_{t-1} &= \frac{1}{2T}\sum_{t=1}^{T}\left(y_t^2 - y_{t-1}^2 - \varepsilon_t^2\right) \\
= & \frac{1}{2T}y_T^2 - \frac{1}{2T}\sum_{t=1}^{T}\varepsilon_t^2 = \frac{1}{2}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_t\right)^2 - \frac{1}{2T}\sum_{t=1}^{T}\varepsilon_t^2 \\
= & \frac{1}{2}\left(\sqrt{T}X_T(1)\right)^2 - \frac{1}{2T}\sum_{t=1}^{T}\varepsilon_t^2 \\
\to & \frac{1}{2}\left(\sigma W(1)\right)^2 - \frac{1}{2}\sigma^2 = \frac{1}{2}\sigma^2\left(W^2(1)-1\right).
\end{aligned}$$

For (4) and (5), since

$$X_T(r) = \begin{cases} 0, & 0 \le r < 1/T \\ y_1/T, & 1/T \le r < 2/T \\ y_2/T, & 2/T \le r < 3/T \\ \vdots & \\ y_{t-1}/T, & (T-1)/T \le r < 1 \\ y_T/T, & r = 1 \end{cases}$$

the integral of $X_T(r)$ in $[0,1]$ is

$$\int_0^1 X_T(r)dr = y_1/T^2 + y_2/T^2 + \cdots + y_{T-1}/T^2 = T^{-2}\sum_{t=1}^{T} y_{t-1},$$

and hence as $T \to \infty$,

$$T^{-3/2}\sum_{t=1}^{T} y_{t-1} = \int_0^1 \sqrt{T}X_T(r)dr \to \int_0^1 \sigma W(r)dr.$$

On the other hand, it follows from

$$
\begin{aligned}
& T^{-3/2}\sum_{t=1}^{T} y_{t-1} \\
= \;& T^{-3/2}[\varepsilon_1 + (\varepsilon_1+\varepsilon_2) + (\varepsilon_1+\varepsilon_2+\varepsilon_3) + \cdots + (\varepsilon_1+\cdots+\varepsilon_{T-1})] \\
= \;& T^{-3/2}[(T-1)\varepsilon_1 + (T-2)\varepsilon_2 + (T-3)\varepsilon_3 + \cdots + (T-(T-1))\varepsilon_{T-1}] \\
= \;& T^{-3/2}\sum_{t=1}^{T}(T-t)\varepsilon_t \\
= \;& T^{-1/2}\sum_{t=1}^{T}\varepsilon_t - T^{-3/2}\sum_{t=1}^{T} t\varepsilon_t
\end{aligned}
$$

that $T^{-3/2}\sum_{t=1}^{T} t\varepsilon_t = T^{-1/2}\sum_{t=1}^{T}\varepsilon_t - T^{-3/2}\sum_{t=1}^{T} y_{t-1} \to \sigma W(1) - \int_0^1 \sigma W(r)dr$ as $T \to \infty$, and hence (4).

For (6), since

$$
\begin{aligned}
ty_{t-1} &= t(\varepsilon_1+\varepsilon_2+\cdots+\varepsilon_{t-1}) \\
&= T\int_{(t-1)/T}^{t/T}([Tr]+1)(\varepsilon_1+\varepsilon_2+\cdots+\varepsilon_{[Tr]})dr \\
&= T^2\int_{(t-1)/T}^{t/T}([Tr]+1)X_T(r)dr,
\end{aligned}
$$

we have

$$
\begin{aligned}
& T^{-5/2}\sum_{t=1}^{T} ty_{t-1} = T^{-1/2}\sum_{t=1}^{T}\int_{(t-1)/T}^{t/T}([Tr]+1)X_T(r)dr \\
= \;& \sum_{t=1}^{T}\int_{(t-1)/T}^{t/T}\frac{[Tr]+1}{T}\sqrt{T}X_T(r)dr = \int_0^1 \frac{[Tr]+1}{T}\sqrt{T}X_T(r)dr \\
\to \;& \sigma\int_0^1 rW(r)dr \text{ as } T \to \infty.
\end{aligned}
$$

(7) is obtained by writing

$$
\begin{aligned}
& T^{-2}\sum_{t=1}^{T} y_{t-1}^2 = T^{-2}\sum_{t=1}^{T}(\varepsilon_1+\varepsilon_2+\cdots+\varepsilon_{t-1})^2 \\
= \;& \sum_{t=1}^{T}\int_{(t-1)/T}^{t/T}\left(\sqrt{T}X_T(r)\right)^2 dr = \int_0^1 \left(\sqrt{T}X_T(r)\right)^2 dr
\end{aligned}
$$

and applying the continuous mapping theorem, where $TX_T(r) \equiv \sum_{t=1}^{[Tr]}\varepsilon_t$ is used. ■

- **Applications to Regression Models:**

Theorem 2 Consider the following three models:

$$\begin{aligned} \text{Model 1} &: \quad y_t = a_1 y_{t-1} + \varepsilon_t, \ a_1 = 1; \\ \text{Model 2} &: \quad y_t = a_0 + a_1 y_{t-1} + \varepsilon_t, \ a_0 = 0, \ a_1 = 1; \\ \text{Model 3} &: \quad y_t = a_0 + a_1 y_{t-1} + \beta t + \varepsilon_t, \ a_0 = 0, \ a_1 = 1, \ \beta = 0. \end{aligned}$$

(i) In Model 1, the data are generated by a random walk: $y_t = y_{t-1} + \varepsilon_t$, but the model is estimated by OLS: $y_t = a_1 y_{t-1} + \varepsilon_t$. The OLS estimator for a_1 satisfies that

$$T(\hat{a}_1 - 1) \to \frac{\int_0^1 W(r)dW(r)}{\int_0^1 W^2(r)dr} = \frac{(W^2(1) - 1)/2}{\int_0^1 W^2(r)dr}$$

where $W(r)$ is standard Brownian motion.

(ii) In Model 2, the data are generated by a random walk: $y_t = y_{t-1} + \varepsilon_t$, but the model is estimated by OLS: $y_t = a_0 + a_1 y_{t-1} + \varepsilon_t$. The OLS estimator for a_1 satisfies that

$$T(\hat{a}_1 - 1) \to \frac{\int_0^1 \bar{W}(r)dW(r)}{\int_0^1 \bar{W}^2(r)dr},$$

where $\bar{W}(r) = W(r) - \int_0^1 W(r)dr$ is the demeaned Brownian motion.

(iii) In Model 3, the data are generated by a random walk: $y_t = y_{t-1} + \varepsilon_t$, but the model is estimated by OLS: $y_t = a_0 + a_1 y_{t-1} + \beta t + \varepsilon_t$. The OLS estimator for a_1 satisfies that

$$T(\hat{a}_1 - 1) \to \frac{\int_0^1 W^*(r)dW(r)}{\int_0^1 W^{*2}(r)dr},$$

where $W^*(r) = W(r) - 4\left(\int_0^1 W(r)dr - \frac{3}{2}\int_0^1 rW(r)dr\right) + 6r\left(\int_0^1 W(r)dr - 2\int_0^1 rW(r)dr\right)$ is the demeaned and detrended Brownian motion.

Proof: (i) For Model 1, the true model is $y_t = y_{t-1} + \varepsilon_t$. Since

$$\hat{a}_1 = \frac{\sum_{t=2}^T y_t y_{t-1}}{\sum_{t=2}^T y_{t-1}^2} = 1 + \frac{\sum_{t=2}^T \varepsilon_t y_{t-1}}{\sum_{t=2}^T y_{t-1}^2},$$

from Theorem 1, we obtain

$$\begin{aligned} T(\hat{a}_1 - 1) &= \frac{T^{-1}\sum_{t=2}^T \varepsilon_t y_{t-1}}{T^{-2}\sum_{t=2}^T y_{t-1}^2} \\ &\to \left(\sigma^2 \int_0^1 W^2(r)dr\right)^{-1} \frac{1}{2}\sigma^2(W^2(1) - 1) \\ &= \left(\int_0^1 W^2(r)dr\right)^{-1} \int_0^1 W(r)dW(r). \end{aligned}$$

(ii) For Model 2, the true model is $y_t = y_{t-1} + \varepsilon_t$. By the OLS estimator in $y_t = a_0 + a_1 y_{t-1} + \varepsilon_t$,

$$\hat{a}_1 = \frac{T \sum_{t=1}^T y_t y_{t-1} - \left(\sum_{t=1}^T y_t\right)\left(\sum_{t=1}^T y_{t-1}\right)}{T \sum_{t=1}^T y_{t-1}^2 - \left(\sum_{t=1}^T y_{t-1}\right)^2},$$

we have

$$\begin{aligned}
T(\hat{a}_1 - 1) &= T \frac{T \sum_{t=1}^T \varepsilon_t y_{t-1} - \left(\sum_{t=1}^T \varepsilon_t\right)\left(\sum_{t=1}^T y_{t-1}\right)}{T \sum_{t=1}^T y_{t-1}^2 - \left(\sum_{t=1}^T y_{t-1}\right)^2} \\
&= \frac{T^{-1} \sum_{t=1}^T \varepsilon_t y_{t-1} - \left(T^{-1/2} \sum_{t=1}^T \varepsilon_t\right)\left(T^{-3/2} \sum_{t=1}^T y_{t-1}\right)}{T^{-2} \sum_{t=1}^T y_{t-1}^2 - \left(T^{-3/2} \sum_{t=1}^T y_{t-1}\right)^2} \\
&\to \frac{\int_0^1 W(r)dW(r) - W(1)\int_0^1 W(r)dr}{\int_0^1 W^2(r)dr - \left(\int_0^1 W(r)dr\right)^2} \\
&= \left(\int_0^1 \bar{W}^2(r)dr\right)^{-1} \int_0^1 \bar{W}(r)dW(r),
\end{aligned}$$

where $\bar{W}(r) \equiv W(r) - \int_0^1 W(r)dr$.

An alternative proof: Let $\bar{y}_t$ be the demeaned y_t, i.e. the residuals from the OLS regression $y_t = \hat{a}_0 + \bar{y}_t$. Denote $\bar{W}(r)$ as the limit distribution of $\frac{1}{\sqrt{T}\sigma}\bar{y}_{[Tr]}$, i.e. $\frac{1}{\sqrt{T}}\bar{y}_{[Tr]} \to \sigma \bar{W}(r)$. It follows from $\bar{y}_t = y_t - \hat{a}_0 = y_t - T^{-1}\sum_{t=1}^T y_t$ that $\bar{y}_t = \bar{y}_{t-1} + \varepsilon_t$ and

$$\frac{1}{\sqrt{T}}\bar{y}_{[Tr]} = \frac{1}{\sqrt{T}} y_{[Tr]} - T^{-3/2}\sum_{t=1}^T y_t \to \sigma\left(W(r) - \int_0^1 W(r)dr\right).$$

Therefore $\bar{W}(r) = W(r) - \int_0^1 W(r)dr$. Then from (i),

$$\begin{aligned}
T\left(\hat{a}_1 - 1\right) &= \frac{T^{-1}\sum_{t=1}^T \varepsilon_t \bar{y}_{t-1}}{T^{-2}\sum_{t=1}^T \bar{y}_{t-1}^2} \\
\to \left(\int_0^1 \bar{W}^2(r)dr\right)^{-1} \int_0^1 \bar{W}(r)d\bar{W}(r) &= \left(\int_0^1 \bar{W}^2(r)dr\right)^{-1} \int_0^1 \bar{W}(r)dW(r).
\end{aligned}$$

(iii) Omitted (see Hamilton(1994)). ■

- **Limit Distribution of t-Statistic for Models 1, 2 and 3:**

Model 1: Under the null: $a_1 = 1$, the OLS estimator $\hat{a}_1$ is a consistent estimator, and hence

$$\hat{\sigma}_T^2 \equiv \frac{1}{T-1} \sum_{t=1}^{T} (y_t - \hat{a}_1 y_{t-1})^2 \to \sigma^2 \text{ in probability.}$$

Then

$$\begin{aligned} t_T &\equiv \frac{\hat{a}_1 - 1}{s.e.(\hat{a}_1)} = \frac{T^{-1}\sum_{t=1}^{T} y_{t-1}\varepsilon_t}{\hat{\sigma}\sqrt{T^{-2}\sum_{t=1}^{T} y_{t-1}^2}} \\ &\to \frac{\frac{1}{2}\sigma^2(W^2(1)-1)}{\sigma\sqrt{\sigma^2\int_0^1 W^2(r)dr}} = \frac{\int_0^1 W(r)dW(r)}{\sqrt{\int_0^1 W^2(r)dr}}. \end{aligned}$$

The t-statistic converge weakly to a functional of the Brownian motion with asymmetric limit distribution.

Model 2: Under the null: $a_1 = 1$,

$$t_T \equiv \frac{\hat{a}_1 - 1}{s.e.(\hat{a}_1)} \to \frac{\int_0^1 \bar{W}(r)dW(r)}{\sqrt{\int_0^1 \bar{W}^2(r)dr}}.$$

Model 3: Under the null: $a_1 = 1$,

$$t_T \equiv \frac{\hat{a}_1 - 1}{s.e.(\hat{a}_1)} \to \frac{\int_0^1 W^*(r)dW(r)}{\sqrt{\int_0^1 W^{*2}(r)dr}}.$$

- **Random Walk with a Drift: What is the Limit Distribution of the OLS Estimator for the autoregression?**

Consider the following random walk plus drift process:

$$y_t = a + \rho y_{t-1} + \varepsilon_t, \tag{8}$$

where $a \neq 0$, $\rho = 1$, and ε_t is i.i.d. Write

$$y_t = y_0 + at + \sum_{i=1}^{t} \varepsilon_i \equiv y_0 + at + \xi_t,$$

which includes a deterministic trend at, where $\xi_t \equiv \sum_{i=1}^{t} \varepsilon_i$. Hence,

$$\sum_{t=1}^{T} y_{t-1} = Ty_0 + \frac{T(T-1)a}{2} + \sum_{t=1}^{T} \xi_{t-1}$$

and as $T \to \infty$, in probability,

$$T^{-2}\sum_{t=1}^{T} y_{t-1} = T^{-1}y_0 + \frac{T(T-1)a}{2T^2} + T^{-1/2}\left(T^{-3/2}\sum_{t=1}^{T}\xi_{t-1}\right) \to \frac{a}{2},$$

which means that $T^{-2}\sum_{t=1}^{T} y_{t-1}$ does not have a non-degenerate distribution duo to the nonzero drift a. Note that

$$\sum_{t=1}^{T} y_{t-1}^2 = \sum_{t=1}^{T}\left(y_0 + a(t-1) + \xi_{t-1}\right)^2$$

$$= Ty_0^2 + a^2\sum_{t=1}^{T}(t-1)^2 + \sum_{t=1}^{T}\xi_{t-1}^2 + 2y_0 a\sum_{t=1}^{T}(t-1) + 2y_0\sum_{t=1}^{T}\xi_{t-1} + 2ay_0\sum_{t=1}^{T}(t-1)\xi_{t-1},$$

$$T^{-3}a^2\sum_{t=1}^{T}(t-1)^2 \to \frac{a^2}{3}, \quad T^{-2}\sum_{t=1}^{T}\xi_{t-1}^2 \to \sigma^2\int_0^1 W^2(r)dr,$$

$$T^{-3/2}\sum_{t=1}^{T}\xi_{t-1} \to \sigma\int_0^1 W(r)dr, \;\; T^{-5/2}\sum_{t=1}^{T}(t-1)\xi_{t-1} \to \sigma\int_0^1 rW(r)dr,$$

we have $T^{-3}\sum_{t=1}^{T} y_{t-1}^2 \to \frac{a^2}{3}$ in probability. Now we derive the limit distribution of the OLS estimators of $\hat{a}$ and $\hat{\rho}$. Since

$$\begin{pmatrix} \hat{a}-a \\ \hat{\rho}-1 \end{pmatrix} = \begin{pmatrix} T & \sum_{t=1}^{T} y_{t-1} \\ \sum_{t=1}^{T} y_{t-1} & \sum_{t=1}^{T} y_{t-1}^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{t=1}^{T}\varepsilon_t \\ \sum_{t=1}^{T} y_{t-1}\varepsilon_t \end{pmatrix},$$

we have

$$\begin{pmatrix} T^{1/2}(\hat{a}-a) \\ T^{3/2}(\hat{\rho}-1) \end{pmatrix} = \begin{pmatrix} T^{1/2} & 0 \\ 0 & T^{3/2} \end{pmatrix}\begin{pmatrix} \hat{a}-a \\ \hat{\rho}-1 \end{pmatrix}$$

$$= \begin{pmatrix} T^{1/2} & 0 \\ 0 & T^{3/2} \end{pmatrix}\begin{pmatrix} T & \sum_{t=1}^{T} y_{t-1} \\ \sum_{t=1}^{T} y_{t-1} & \sum_{t=1}^{T} y_{t-1}^2 \end{pmatrix}^{-1}\begin{pmatrix} T^{1/2} & 0 \\ 0 & T^{3/2} \end{pmatrix}$$

$$\cdot\begin{pmatrix} T^{-1/2} & 0 \\ 0 & T^{-3/2} \end{pmatrix}\begin{pmatrix} \sum_{t=1}^{T}\varepsilon_t \\ \sum_{t=1}^{T} y_{t-1}\varepsilon_t \end{pmatrix}$$

$$= \left[\begin{pmatrix} T^{-1/2} & 0 \\ 0 & T^{-3/2} \end{pmatrix}\begin{pmatrix} T & \sum_{t=1}^{T} y_{t-1} \\ \sum_{t=1}^{T} y_{t-1} & \sum_{t=1}^{T} y_{t-1}^2 \end{pmatrix}\begin{pmatrix} T^{-1/2} & 0 \\ 0 & T^{-3/2} \end{pmatrix}\right]^{-1}$$

$$\cdot\begin{pmatrix} T^{-1/2}\sum_{t=1}^{T}\varepsilon_t \\ T^{-3/2}\sum_{t=1}^{T} y_{t-1}\varepsilon_t \end{pmatrix}$$

$$= \begin{pmatrix} 1 & T^{-2}\sum_{t=1}^{T} y_{t-1} \\ T^{-2}\sum_{t=1}^{T} y_{t-1} & T^{-3}\sum_{t=1}^{T} y_{t-1}^2 \end{pmatrix}^{-1}\begin{pmatrix} T^{-1/2}\sum_{t=1}^{T}\varepsilon_t \\ T^{-3/2}\sum_{t=1}^{T} y_{t-1}\varepsilon_t \end{pmatrix}.$$

Since

$$\begin{pmatrix} 1 & T^{-2}\sum_{t=1}^{T} y_{t-1} \\ T^{-2}\sum_{t=1}^{T} y_{t-1} & T^{-3}\sum_{t=1}^{T} y_{t-1}^2 \end{pmatrix} \to \begin{pmatrix} 1 & \frac{a}{2} \\ \frac{a}{2} & \frac{a^2}{3} \end{pmatrix}$$

and

$$\begin{aligned} \begin{pmatrix} T^{-1/2}\sum_{t=1}^{T} \varepsilon_t \\ T^{-3/2}\sum_{t=1}^{T} y_{t-1}\varepsilon_t \end{pmatrix} &= \begin{pmatrix} T^{-1/2}\sum_{t=1}^{T} \varepsilon_t \\ T^{-3/2}\sum_{t=1}^{T} \left(y_0 + a(t-1) + \xi_{t-1}\right)\varepsilon_t \end{pmatrix} \\ = & \begin{pmatrix} T^{-1/2}\sum_{t=1}^{T} \varepsilon_t \\ T^{-3/2}\sum_{t=1}^{T} a(t-1)\varepsilon_t \end{pmatrix} + \begin{pmatrix} 0 \\ T^{-3/2}\sum_{t=1}^{T} \left(y_0 + \xi_{t-1}\right)\varepsilon_t \end{pmatrix} \\ \to & N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \sigma^2 \begin{pmatrix} 1 & \frac{a}{2} \\ \frac{a}{2} & \frac{a^2}{3} \end{pmatrix}\right], \end{aligned}$$

we have

$$\begin{aligned} \begin{pmatrix} T^{1/2}\left(\hat{a} - a\right) \\ T^{3/2}\left(\hat{\rho} - 1\right) \end{pmatrix} &\to N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \sigma^2 \begin{pmatrix} 1 & \frac{a}{2} \\ \frac{a}{2} & \frac{a^2}{3} \end{pmatrix}^{-1}\right]. \qquad (9) \\ &= N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \sigma^2 \begin{pmatrix} 4 & -\frac{6}{a} \\ -\frac{6}{a} & \frac{12}{a^2} \end{pmatrix}\right] \end{aligned}$$

Note: Compared with the random walk without a drift, the limt distribution of $T^{1/2}\left(\hat{a} - a\right)$ and $T^{3/2}\left(\hat{\rho} - 1\right)$ have following properties:

(i) Asymptotic Normality. (For the random walk without a drift, the distribution is nonstandard and asymmetric)

(ii) Consistency with the speed rates $T^{-1/2}$ and $T^{-3/2}$, respectively. (For the random walk without a drift, the speed rate of $\hat{\rho}$ is T^{-1}).

- **Example:** Suppose that the true model is a unit root process with a constant

$$y_t = \alpha + y_{t-1} + u_t, \qquad (10)$$

where u_t are i.i.d.$(0, \sigma^2)$. However, we use **a Trend-Stationary Model**

$$y_t = c + \beta t + v_t \qquad (11)$$

for the estimation of β. What is the limit distribution of the OLS estimator $\hat{\beta}$? What happens to the conventional t-test for $\beta = 0$ in Model (11)?

Denote $s_t \equiv \sum_{i=1}^{t} u_i$. Since $y_t = at + s_t$, we have

$$\begin{aligned}
\hat{\beta} &= \frac{T\sum_{t=1}^{T} ty_t - \left(\sum_{t=1}^{T} t\right)\left(\sum_{t=1}^{T} y_t\right)}{T\sum_{t=1}^{T} t^2 - \left(\sum_{t=1}^{T} t\right)^2} \\
&= \frac{T\sum_{t=1}^{T} t\left(at + s_t\right) - \left(\sum_{t=1}^{T} t\right)\left(\sum_{t=1}^{T}\left(at + s_t\right)\right)}{T\sum_{t=1}^{T} t^2 - \left(\sum_{t=1}^{T} t\right)^2} \\
&= a + \frac{T\sum_{t=1}^{T} ts_t - \left(\sum_{t=1}^{T} t\right)\left(\sum_{t=1}^{T} s_t\right)}{T\sum_{t=1}^{T} t^2 - \left(\sum_{t=1}^{T} t\right)^2}.
\end{aligned}$$

Write

$$\begin{aligned}
\sqrt{T}(\hat{\beta} - a) &= \sqrt{T}\frac{T\sum_{t=1}^{T} ts_t - \left(\sum_{t=1}^{T} t\right)\left(\sum_{t=1}^{T} s_t\right)}{T\sum_{t=1}^{T} t^2 - \left(\sum_{t=1}^{T} t\right)^2} \\
&= \frac{T^{-7/2}\left[T\sum_{t=1}^{T} ts_t - \left(\sum_{t=1}^{T} t\right)\left(\sum_{t=1}^{T} s_t\right)\right]}{T^{-4}\left[T\sum_{t=1}^{T} t^2 - \left(\sum_{t=1}^{T} t\right)^2\right]}.
\end{aligned}$$

Since $lim_{T\to\infty}T^{-4}\left[T\sum_{t=1}^{T} t^2 - \left(\sum_{t=1}^{T} t\right)^2\right] = 1/12$ and

$$\begin{aligned}
& T^{-5/2}\sum_{t=1}^{T} ts_t - \left(T^{-2}\sum_{t=1}^{T} t\right)\left(T^{-3/2}\sum_{t=1}^{T} s_t\right) \\
& \to \sigma\int_0^1 rW(r)dr - \frac{1}{2}\sigma\int_0^1 W(r)dr,
\end{aligned}$$

we have,

$$\sqrt{T}(\hat{\beta} - a) \to 12\sigma\int_0^1 \left(r - \frac{1}{2}\right) W(r)dr$$

and $\hat{\beta} \to a$ in probability. Now it is concluded that, as long as $a \neq 0$, the OLS estimator $\hat{\beta}$ from the model (11) has non-zero probability limit, and hence the t-test of the null: $a = 0$ would be significant. It seems that we would never reject the deterministic trend in a difference stationary time series. The significance of the t-test of $\hat{\beta}$ and $\hat{c}$ in the model (11) can not allow one to distinguish (10) and (11). The first thing we should do is to conduct the unit root test for the series; only when the unit root test is rejected can we conduct the conventional t-test for the parameters c and β in the model (11).

- **Unit Root Test:**

1) Dickey-Fuller (DF) Test: The three basic models used for regression

$$\begin{aligned} &\text{Model 1:} && y_t = \rho y_{t-1} + \varepsilon_t, \\ &\text{Model 2:} && y_t = \alpha + \rho y_{t-1} + \varepsilon_t, \\ &\text{Model 3:} && y_t = \alpha + \delta t + \rho y_{t-1} + \varepsilon_t. \end{aligned}$$

can be equivalently written, respectively, as

$$\text{Model 1:} \quad \Delta y_t = \gamma y_{t-1} + \varepsilon_t, \tag{12}$$

$$\text{Model 2:} \quad \Delta y_t = \alpha + \gamma y_{t-1} + \varepsilon_t, \tag{13}$$

$$\text{Model 3:} \quad \Delta y_t = \alpha + \delta t + \gamma y_{t-1} + \varepsilon_t \tag{14}$$

where $\gamma = \rho - 1$. The null of unit root is $\gamma = 0$. Denote the test statistics for Model 1, 2, and 3, respectively, τ, τ_μ, and τ_τ:

$$\tau \text{ or } \tau_\mu \text{ or } \tau_\tau = \frac{\hat{\rho} - 1}{\text{s.e.}\,(\hat{\rho})} = \frac{\hat{\gamma} - 0}{\text{s.e.}\,(\hat{\gamma})}$$

Dickey and Fuller (1976) apply the distribution of τ-statistic to test for the unit root $\rho = 1$, i.e. $\gamma = 0$. The critical values of the τ-statistics for the three models above are different. For example, the critical values for 99% and 95% confidence intervals for $T = 100$ are, respectively,

Model used for regression	1% critical value	5% critical value
1: no constant or time trend	−2.60	−1.95
2: no time trend	−3.51	−2.89
3: constant or time trend	−4.04	−3.45

Monte Carlo Experiments (simulation) has been used to determine the critical value of the test. Evans and Savin (1981) show that DF test is equivalent to Wald or LM test under the normality assumption of the error term.

Further, Dickey and Fuller (1981) provide three additional F-statistics, called ϕ_1, ϕ_2 and ϕ_3, to test the joint hypothses on the coefficients in the applied regression model. They are defined as the ordinary F-tests:

$$\phi_i = \frac{\left(SSR_{restricted} - SSR_{unrestricted}\right)/r}{SSR_{unrestricted}/(T-k)}.$$

For model 2, the test statistic for the joint hypothesis $\alpha = \gamma = 0$ is ϕ_1;

For model 3, the test statistic for the joint hypothesis $\alpha = \delta = \gamma = 0$ is ϕ_2;

For model 3, the test statistic for the joint hypothesis $\delta = \gamma = 0$ is ϕ_3.

Summarily,

Model used for regression	Null Hypothesis H_0	Test Statistic
Model 3: $\Delta y_t = \alpha + \delta t + \gamma y_{t-1} + \varepsilon_t$	$\gamma = 0$	τ_τ
	$\gamma = \delta = 0$	ϕ_3
	$\alpha = \gamma = \delta = 0$	ϕ_2
Model 2: $\Delta y_t = \alpha + \gamma y_{t-1} + \varepsilon_t$	$\gamma = 0$	τ_μ
	$\alpha = \gamma = 0$	ϕ_1
Model 1: $y_t = \rho y_{t-1} + \varepsilon_t$,	$\gamma = 0$	τ

Note: Problems:
(i) For y_t, AR(p)?
(ii) For the error term, MA(q)?
(iii) When to include the constant or time trend in the regression model?
(iv) More than one unit roots?
(v) Structural change?

2) **Augmented Dickey-Fuller (ADF) Test (Said and Dickey test):** Extend the DF test to ARMA(p,q) models

$$y_t = a_1 y_{t-1} + \cdots + a_p y_{t-p} + \varepsilon_t + b_1 \varepsilon_{t-1} + \cdots + b_q \varepsilon_{t-q}$$

or

$$(1 - A(L))y_t \equiv (1 - a_1 L - \cdots - a_p L^p) y_t = (1 + b_1 L + \cdots + b_q L^q)\varepsilon_t \equiv B(L)\varepsilon_t,$$

where all the roots of the equation $B(r) = 0$ lie outside the unit circle (i.e. the invertibility condition of ARMA model is satisfied), and $1 - A(L) = 0$ may have one positive unit root and the rest outside the unit circle. We are interested in testing for the presence of a unit root in the characteristic equation $1 - A(L) = 0$. Suppose $B(L)^{-1} = 1 - C(L)$. Since

$$B(L)^{-1}(1 - A(L))y_t = \varepsilon_t$$

or

$$(1 - C(L))(1 - A(L))y_t = \varepsilon_t,$$

we have

$$y_t = L^{-1}\left[A(L) + C(L) - A(L)C(L)\right] y_{t-1} + \varepsilon_t \equiv \Omega(L) y_{t-1} + \varepsilon_t$$

or

$$\Delta y_t = (\Omega(L) - 1)y_{t-1} + \varepsilon_t$$

Applying the Beveridge-Nelson decomposition

$$\Omega(z) = \sum_{j=0}^{\infty} w_j z^j = \Omega(1) - (1 - z)\tilde{\Omega}(z)$$

where $\tilde{\Omega}(z) = \sum_{j=0}^{\infty} \tilde{w}_j z^j$ and $\tilde{w}_j = \sum_{k=j+1}^{\infty} w_k$, we have

$$\begin{aligned}
\Delta y_t &= \left[\Omega(1) - (1 - L)\tilde{\Omega}(L) - 1\right] y_{t-1} + \varepsilon_t \\
&= [\Omega(1) - 1]\, y_{t-1} - \tilde{\Omega}(L)\Delta y_{t-1} + \varepsilon_t \\
&= \gamma y_{t-1} - \sum_{i=0}^{\infty} \tilde{w}_i \Delta y_{t-i-1} + \varepsilon_t \\
&= \gamma y_{t-1} + \sum_{i=2}^{\infty} \beta_i \Delta y_{t-i+1} + \varepsilon_t,
\end{aligned}$$

where $\gamma = \Omega(1) - 1$. **(Note: A simpler derivation for AR(p) models is given on page 190 in Enders's book)**. This contains infinite number of regressors and is inadequate for estimation. Thus, we use its **truncated version**

$$\text{Model } 1'\text{: } \Delta y_t = \gamma y_{t-1} + \sum_{i=2}^{k} \beta_i \Delta y_{t-i+1} + u_{kt},$$

where k is chosen as a finite number (often by AIC or SBC). Said and Dickey (1984) derive the limiting distribution of the t-test (ADF test) for the null hypothesis $\gamma = 0$, which is the same as that in Model 1 above. Therefore, the test statistic for $\gamma = 0$ is the same as the τ-statistic above. Similarly, the tests for $\gamma = 0$ in the following models

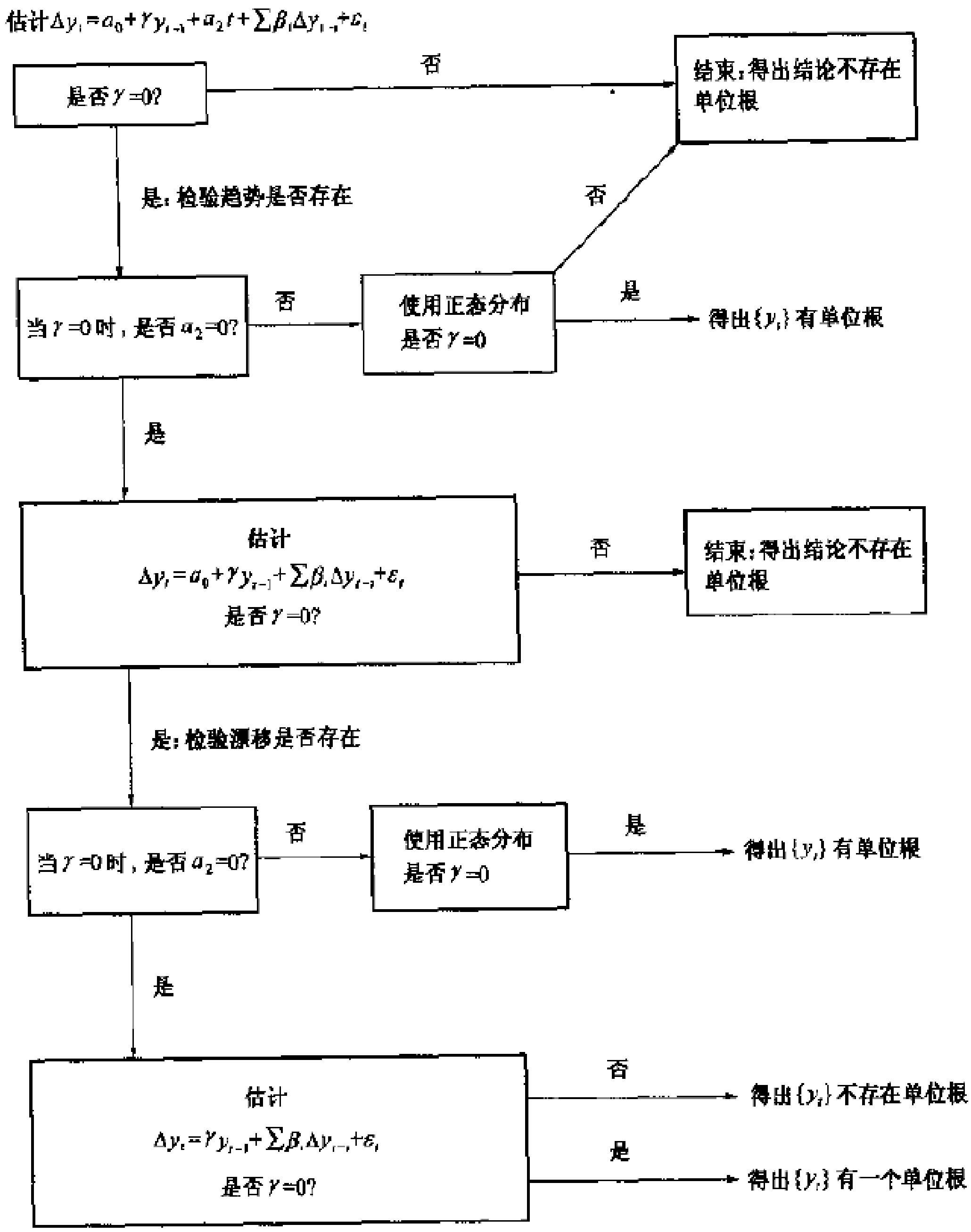
$$\text{Model } 2'\text{: } \Delta y_t = \alpha + \gamma y_{t-1} + \sum_{i=2}^{k} \beta_i \Delta y_{t-i+1} + u_{kt}$$

and

$$\text{Model } 3'\text{: } \Delta y_t = \alpha + \gamma y_{t-1} + \delta t + \sum_{i=2}^{k} \beta_i \Delta y_{t-i+1} + u_{kt}$$

are also the same as those in Model 2 and Model 3, respectively.

单位根检验的一般步骤



估计$\Delta y_t=a_0+\gamma y_{t-1}+a_2t+\sum\beta_i\Delta y_{t-i}+\varepsilon_t$
是否γ=0?
否
结束：得出结论不存在单位根
是：检验趋势是否存在
当γ=0时，是否a_2=0?
否
使用正态分布是否γ=0
否
是
得出$\{y_t\}$有单位根
是
估计
$\Delta y_t=a_0+\gamma y_{t-1}+\sum\beta_i\Delta y_{t-i}+\varepsilon_t$
是否γ=0?
否
结束：得出结论不存在单位根
是：检验漂移是否存在
当γ=0时，是否a_2=0?
否
使用正态分布是否γ=0
是
得出$\{y_t\}$有单位根
是
估计
$\Delta y_t=\gamma y_{t-1}+\sum\beta_i\Delta y_{t-i}+\varepsilon_t$
是否γ=0?
否
得出$\{y_t\}$不存在单位根
是
得出$\{y_t\}$有一个单位根

- **A Four-step Test Approach: when the form of the DGP is completely unknown.** See Enders's book: P213 **Figure 4.13**. Note that 1) Plotting the data is usually an important indicator of presence of deterministic regressors; 2) Theoretical consideration might suggest the appropriate deterministic regressors. See GDP and Unit Roots example on P214.

- **An Empirical Example: PPP (Purchasing Power Parity) Puzzle:** Purchasing power parity does not significantly hold in empirical studies. Why? Academic and Empirical explanations.

- **Some Problems about the Unit Root Test:**

1. **Choice of** k derived from an $ARMA(p,q)$ model or p in the $AR(p)$ model——use t-test and F-test (Note that the OLS estimator for the coefficient β is asymptotically normally distributed and hence the t-test and F-test are appropriate (see Sims, Stock and Watson (1990)), but, since the estimator for γ is not normally distributed, we have to apply the DF test for unit root).

2. **Multiple unit roots**——If two unit roots are suspected, estimate

$$\Delta^2 y_t = a_0 + \beta_1 \Delta y_{t-1} + \varepsilon_t$$

and use the appropriate statistic ($\tau, \tau_\mu, \tau_\tau$ depending on the deterministic elements actually included in the regression) to test the null $H_0 : \beta_1 = 0$. If accepting H_0, conclude that $\{y_t\} \sim I(2)$; otherwise, go on to test a single unit root by estimating

$$\Delta^2 y_t = a_0 + \beta_1 \Delta y_{t-1} + \beta_2 y_{t-1} + \varepsilon_t$$

and testing the null $\beta_2 = 0$. If accepting, conclude that $\{y_t\}$ has a single unit root; otherwise, rejecting the null deduces that $\{y_t\}$ is stationary. Similarly for the test of more than two unit roots.

3. **Seasonal unit roots**——Note that, the first difference of a seasonal unit root processes will not be stationary; however, the seasonal difference of a unit root process may be stationary.(Study the process $\{y_t\} : y_t = y_{t-4} + \varepsilon_t$). Some methods to treat seasonality:

(a) Introduce seasonal dummy variables, e.g. for quarterly data,

$$\Delta y_t = a_0 + \alpha_1 D_1 + \alpha_2 D_2 + \alpha_3 D_3 + \gamma y_{t-1} + \sum_{i=2}^{p} \beta_i \Delta y_{t-i+1} + \varepsilon_t.$$

Use the DF τ_μ statistic to test the null $H_0 : \gamma = 0$. If rejecting H_0, conclude that $\{y_t\}$ is stationary. For a time trend in the model, use the τ_τ statistic to test H_0. (The shortcoming: hard to test the hypothesis about a_0.Solution is to introduce centered seasonal dummy variables s.t. the mean for each of the dummy variables is zero).

(b) Test for seasonality: e.g. for quarterly data, if $\gamma = 1$,

$$(1 - \gamma L^4)y_t = (1 - \gamma^{1/4}L)(1 + \gamma^{1/4}L)(1 - i\gamma^{1/4}L)(1 + i\gamma^{1/4}L)y_t = \varepsilon_t$$

only allows for a unit root at an annual frequency, i.e. a seasonal unit root. Generally, consider

$$(1 - a_1 L)(1 + a_2 L)(1 - i a_3 L)(1 + i a_4 L) = \varepsilon_t. \tag{15}$$

(a) If $a_1 = a_2 = a_3 = a_4 = 1$, there is a seasonal unit root ($\gamma = 1$); (b) If $a_1 = 1$, $y_t = y_{t-1}$ is one homogeneous solution of the model; there is no seasonal unit root. (Differencing the data is appropriate); (c) If $a_2 = 1$, $y_t + y_{t-1} = 0$ is one homogeneous solution of the model; there is a semiannual unit root; (d) If a_3 or $a_4 = 1$, $y_t = i y_{t-1}$ or $y_t = -i y_{t-1}$ is the homogeneous solution of the model; there is a seasonal unit root. Use the difference $\Delta_4 y_t = (1 - L^4)y_t$. Generally, by a Taylor expansion at $a_1 = a_2 = a_3 = a_4 = 1$, the model (15) above is approximated by

$$\begin{aligned} (1 - L^4)y_t &= \gamma_1(1 + L + L^2 + L^3)y_{t-1} - \gamma_2(1 - L + L^2 - L^3)y_{t-1} \\ &\quad + \gamma_5(1 - L^2)y_{t-1} - \gamma_6(1 - L^2)L y_{t-1} + \varepsilon_t \\ &\equiv \gamma_1 y_{1t-1} - \gamma_2 y_{2t-1} + \gamma_5 y_{3t-1} - \gamma_6 y_{3t-2} + \varepsilon_t \end{aligned} \tag{16}$$

where $\gamma_1 = a_1 - 1, \gamma_2 = a_2 - 1, \gamma_5 = (a_3 - a_4)i, \gamma_6 = a_3 + a_4 - 2$, and

$$\begin{aligned} y_{1t-1} &= (1 + L + L^2 + L^3)y_{t-1}, \\ y_{2t-1} &= (1 - L + L^2 - L^3)y_{t-1}, \\ y_{3t-1} &= (1 - L^2)y_{t-1}, \\ y_{3t-2} &= (1 - L^2)L y_{t-1}. \end{aligned}$$

In application, for quarterly data $\{y_t\}$, estimate the model (16) (when necessary, modify the form by including the intercept, deterministic seasonal dummies, and a linear time trend), insuring that the residuals approximate a white-noise process. **First, test** $\gamma_1 = 0$ by using the critical values reported in Hylleberg, et al (1990) (see also P.199 in the textbook). If accepting the null, conclude that $a_1 = 1$ and there is a nonseasonal unit root. **Then t-test** $\gamma_2 = 0$**.** If accepting $\gamma_2 = 0$, conclude that $a_2 = 1$ and there is a unit root with a semiannual frequency. **Finally, F-test** $\gamma_5 = \gamma_6 = 0$. If the calculated value is less than the critical value reported in Hylleberg, et al (1990), conclude that $\gamma_5 = \gamma_6 = 0$ and there is a seasonal unit root.

4. **Structural Unit Root Test:** Perron's Test**.**

5. **KPSS Test:** The null is that the series is stationary.